Cover and Title Page

Final Performance Report for Project

Architecture and Performance Analysis of General Bio-Molecular Networks

Contract/Grant #: FA9550-10-1-0128

Table of Contents

REPORT DOCUMENTATION PAGE

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
14-01-2012	Final Performance Report	15-04-2010 TO 14-10-2011

4. TITLE AND SUBTITLE
Architecture and Performance Analysis of Bio-Molecular Network

5a. CONTRACT NUMBER
FA9550-10-1-0128

5b. GRANT NUMBER
AFOSR/PKR1

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Li, Wei, W.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Texas Southern University, 3100 Cleburne ST, Houston, TX 77004

8. PERFORMING ORGANIZATION REPORT NUMBER
3ZFG0

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research

875 North Randolph St.

Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-OSR-VA-TR-2012-0776

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A

13. SUPPLEMENTARY NOTES
This project provided a detailed framework to describe the components, reactions, and internal and external connections of the cellular control systems, developed a stochastic model of the generalized biomolecular network.

14. ABSTRACT
This project provided a detailed framework to describe the components, reactions, and internal and external connections of the cellular control systems, developed a stochastic model of the generalized biomolecular network and obtained an analytic approach for the solution of the stochastic molecular network. Specifically, we have (1) Derived a novel and explicit iterative method (both the forward iterative method and backward iterative method) to calculate the solution of the Network Chemical Master Equation (CME); (2) Completed the numerical analysis based on our novel method and simulation analysis based on the nature of the molecular network for an example; (3) The result from numerical analysis and simulation analysis is almost exact same, which confirmed the correction of our theoretical result; and (4) The time to calculate the solution of CME by using our method is about 10 times faster than the time to calculate it by using the standard method. Our method is expected to be much better, in terms of the running time, for the system with more molecules.

15. SUBJECT TERMS
Stochastic Bio-molecular Networks, Explicit Iterative Method, Network Chemical Master Equation.

16. SECURITY CLASSIFICATION OF: U			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Dr. Li, Wei W.
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU		**19b. TELEPHONE NUMBER** (include area code) 713-313-1871

2

3. Report Documentation

3.1 Executive summary

The research on this project developed a detailed framework to describe the components, reactions, and internal and external connections of cellular control systems. Based on this framework, the principal investigator (PI) built a stochastic model of the generalized biomolecular network and developed an analytic approach for the solution. Furthermore, the PI made two major contributions as below:

1. The PI derived a novel iterative method (using both the forward and backward iterative method) to calculate the solution of the Network Chemical Master Equation (CME). By using a simple reaction network, the isomerization reaction, the PI completed the numerical analysis based on both novel analytical method and direct simulation analysis for the test reaction network.

2. The PI proposed a novel model of general bio-molecular networks, where a number of different classes of the molecules are involved and among each class there is a number of different species of molecules. He also introduced the concept of the eventual reaction time (ERT) connecting one molecular species to another in this general network. Furthermore, the PI derived the mathematical formulas for the probability density function of the eventual reaction time, the *k-th* moment of this eventual reaction time, the probability that reaction *j* triggers the ERT event and the joint density function that reaction *j* triggers the ERT event at time *t*.

The results from numerical analysis and simulation analysis were conducted and were verified to be almost exactly same, which supports the validity of the theoretical result. Importantly, the time to calculate the solution of CME and ERT using our method is about 10 times faster than the time using the standard methods. There two novel methods are expected to be superior, in terms of the running time, for more complex systems.

3.2 List of people involved

The total involved person in this research were:

Dr. Wei Wayne Li	**PI of the Project** Professor Department of Computer Science Texas Southern University, Houston, TX E-mail: liw@tsu.edu
Dr. John M. Frazier	**Technical Advisor of the Project** ST - Senior Scientist Human Effectiveness Directorate Air Force Research Laboratory Wright-Patterson Air Force Base E-mail: john.frazier@wpafb.af.mil
Mr. Sovandara Chea	**Master Student partially supported from the Project** Master student Department of Computer Science Texas Southern University, Houston, TX E-mail: cheas@tsu.edu
Mr. Zhao Zhang	**Master Student partially supported from the Project** Master student Department of Computer Science Texas Southern University, Houston, TX E-mail: z.zhao1124@student.tsu.edu

3.3   Major Contribution

In all organisms, networks of coupled biochemical reactions and feedback signals organize developmental pathways, metabolic pathways, and progression through the cell cycle. Within these regulatory networks, genetic activity is controlled by molecular signals that determine when and how often a given gene is transcribed. The overall purpose of this research project is to develop a new mathematical description of the biomolecular networks that control cellular behavior, formulate both novel analytic and computational tools to investigate the behavior of these networks and apply these approaches to evaluate the performance of engineered bioconstructs. The major contribution of this project included

➢ Developed a general biomolecular network description of cellular control systems from gene to output function and provided a detailed framework to describe the components, reactions, and internal and external connections of these systems.

➢ Built a stochastic model of the generalized biomolecular network and developed an analytic approach for the time-dependent solution and investigate computational methods to approximate these solutions.

➢ Successfully applied the stochastic model and analytical methods to answer specific questions about system behavior, including
  ✓ Define the "distance" between two nodes in terms of the time it takes for a variation introduced at the first node to propagate to the second node. Investigate the hysteresis effect upon reversal of the nodes.
  ✓ Compute the distribution of time intervals from a given current state to a defined final state and identify the most likely reaction pathway that makes this happen.

A more detailed report is included in the following two sections:

3.3.1 A General Bio-Molecular Network Model and the Solution of the Network Chemical Master Equation

The research of this research effort has developed a modeling framework that can be used to investigate complex biomolecular network models. Specifically, this research has reached the following goals
  ➢ Understood the molecular basis of complex stochastic biomolecular processes,
  ➢ Developed a stochastic model to describe these processes and interactions,
  ➢ Derived an exact solution of the Chemical Master Equation for the general bio-molecular network and expressed the result as a series representation in terms of the propensity functions.
  ➢ Simulated biomolecular events during the operation of biological constructs under different experimental conditions.

Final Performance Report

At first, the PI constructed and defined architecture of biomolecular network systems as depicted in Figure 1 below.
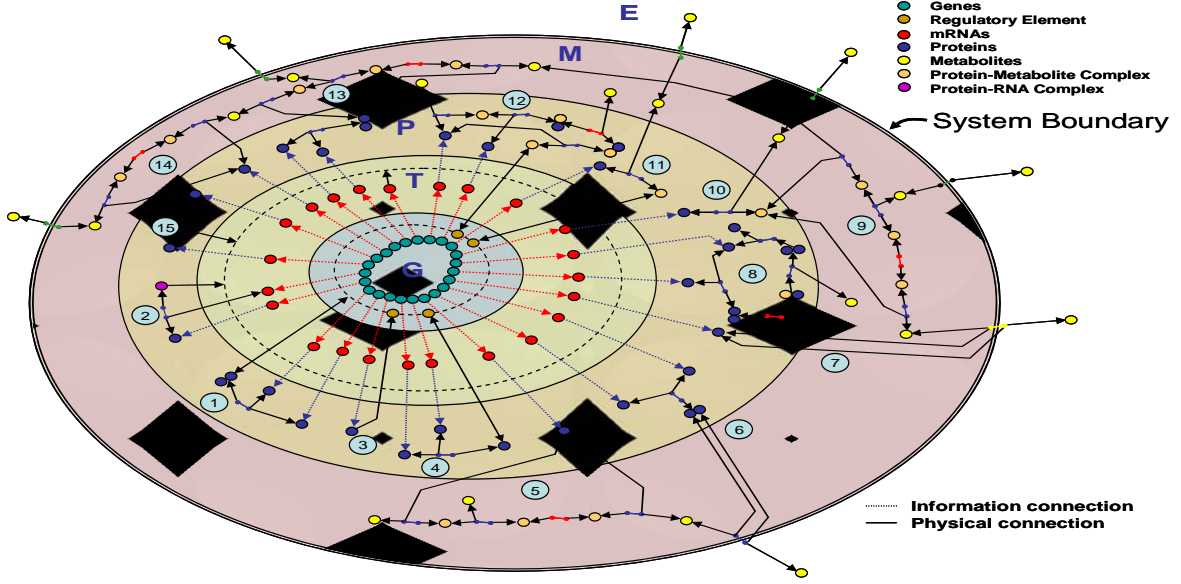

Figure 1. Detailed architecture and connections between system components in our model of a stochastic biomolecular network.

Based on this description, The PI constructed a multi-dimensional Markov Jump Process on the state space

$$\Omega = \left\{ \boldsymbol{x} \equiv \left(x_{1,1}, x_{1,2}, \cdots, x_{1,N_1}; x_{2,1}, x_{2,2}, \cdots, x_{2,N_2}; \cdots; x_{10,1}, x_{10,2}, \cdots, x_{10,N_{10}} \right) \mid 0 \le x_{i,j} \le N_{i,j} \right\} \qquad (1)$$

in which $i = 1, 2, ..., 10$ and $j = 1, 2, ..., N_i$. Here it is assumeed there are a total of *M* **fundamental chemical reactions** in the network, say $R_1, R_2, \cdots, R_M$. Each reaction changes the population of at least one species of molecules and a population change of any species comes from a reaction. Mathematically speaking, reaction R_k is characterized by a **reaction probability constant** c_k, which is assumed to be independent of time *t*. The change in numbers of molecules of the biomolecular network is described by a $N \equiv \sum_{i=1}^{10} N_i$ dimensional vector, called the **state-change vector**, $\boldsymbol{v}_k = (\boldsymbol{v}_{1,k}, \boldsymbol{v}_{2,k}, \cdots, \boldsymbol{v}_{10,k})$, where $\boldsymbol{v}_{i,k} = (v_{i,1,k}, v_{i,2,k}, \cdots, v_{i,N_i,k})$, and $v_{i,j,k}$ is the change in the number of (*i,j*)-molecule produced by reaction R_k, for all $i = 1, 2, ..., 10$, $j = 1, 2, ..., N_i$ and $k = 1, 2, ..., M$.

In general, the propensity functions and state-change vectors together will completely characterize all reaction channels. Mathematically speaking, they also characterize the stochastic process **X(t)** as follows

$$P[X(t+dt) = x + v_k / X(t) = x] = a_k(x)dt \ . \tag{2}$$

Now, for any two states $x, y \in \Omega$, we introduce a new function of (x,y) as follows:

$$q_{x,y} = \begin{cases} a_k(x), & \text{if there is a reaction } R_k \text{ such that } y = x + v_k, \\ -a_0(x), & \text{if } y = x, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where $a_0(x) \equiv \sum_{k=1}^{M} a_k(x)$ for any $x \in \Omega$. Based on above assumptions for the reaction rate and

Equation (1), it can be verified that X(t) is a Markov Jump Process and is defined by the

transition rate matrix $Q = (q_{x,y})$, with size of $\mathcal{N} = \prod_{i=1}^{10} \prod_{j=1}^{N_i} (N_{i,j} + 1)$. Let $P(x,t)$ be the probability

that the system is in state x at time $t,$ and $P(t)$ a probability row vector with $P(x,t)$ as element
with the **x**s ordered lexicographically. From the theory of stochastic processes, we have

$$\frac{d}{dt} P(t) = P(t)Q, \quad \text{with initial condition P(0)} = p_0, \tag{4}$$

where P(0) is an **N** –dimensional row vector determined by the initial state of the biomolecular
network. The exact solution for Equation (4) is $P(t) = p_0 e^{Qt}$. Therefore

$$P(x,t) = P(t)e_x = p_0 e^{Qt} e_x, \tag{5}$$

where e_x is a **N** –dimensional column vector with one in the position of vector x and zeroes
otherwise. Furthermore, we will have the following main result. The major result here is that
the solution of the Chemical Master Equation is given by

$$P(x,t) = b_0(x^{(0)}, x) + \sum_{n=1}^{\infty} b_n(x^{(0)}, x) \frac{t^n}{n!}, \tag{6}$$

where $b_n(x^{(0)}, x)$ is given in terms of the propensity functions and state change vectors as:

$$b_0(y^{(0)}, x) = \begin{cases} 1, & \text{if } x = y^{(0)}; \\ 0, & \text{otherwise.} \end{cases}$$

$$b_n(y^{(0)}, x) = -a_0(y^{(0)})b_{n-1}(y^{(0)}, x) + \sum_{k=1}^{M} a_k(y^{(0)})b_{n-1}(y^{(0)} + v_k, x), \quad \text{for n} = 1,2,\cdots, \tag{7}$$

where $y^{(0)}$ is any state in the state space Ω.

3.3.2. Eventual Reaction Time in General Bio-Molecular Networks

The PI proposed a novel model of general bio-molecular networks, where a number of different classes of the molecules are involved and among each class there is a number of different species of molecules. The major contribution in this part included:

- Introduced the concept of the *eventual reaction time (ERT)* connecting one molecular species to another in this general network;
- Derived the mathematical formulas for the probability density function of the eventual reaction time, the *k-th* moment of this eventual reaction time, the probability that reaction *j* triggers the ERT event and the joint density function that reaction *j* triggers the ERT event at time *t*.

At first, the PI considered this ERT problem from the viewpoint of the cell network. Since there are many other types and classes of the molecules at time t_0, we may assume that at t_0 there are also a specific number $x_{p_0,q_0}(t_0)$ of (p_0,q_0)-molecules, in which $p_0 = 1, 2, \cdots, 10;\ q_0 = 1, 2, \cdots, N_{q_0}$ where $(p_0,q_0) \neq (i_0, j_0)$. The set of all of these values taken together describe the state of the system at time t_0. We denote by x_0 this network state at time t_0, which is a specific state of the state space Ω (Equation (3)), and is called the initial state of the network. Since there are many other species from different classes of the molecules at any specific time, at the time when the system **first** attains a specific number x_{i_1,j_1} of (i_1,j_1)-molecules in the network, it is possible to have a number x_{p_1,q_1} of (p_1,q_1)-molecules, in which $p_1 = 1, 2, \cdots, 10;\ q_1 = 1, 2, \cdots, N_{q_1}$ and $(p_1,q_1) \neq (i_1, j_1)$. Any of these states could describe a legitimate destination state. Since, we are only interested in the first time the system contains exactly x_{i_1,j_1} (i_1,j_1)-molecules, we may form a sub-set of the state space, Ω^D, consisting of all those states. For example, if $(i_1, j_1) = (1,2)$, then

$$\Omega^D = \left\{ x \equiv \left(x_{1,1}, x_{i_1,j_1}, \cdots, x_{1,N_1}; x_{2,1}, x_{2,2}, \cdots, x_{2,N_2}; \cdots; x_{10,1}, x_{10,2}, \cdots, x_{10,N_{10}}\right) \in \Omega \right\}.$$

This set is called the set of destination states. If we denote by Ω^E the set of all other state, then $\Omega = \Omega^E + \Omega^D$. Accordingly, we use c_k^E, a_k^E, R_k^E, M^E to denote the reaction probability constant for reaction *k*, the corresponding propensity function, the reactions and the total reaction numbers in Ω^E. Based on above two steps description, if we denote by $\tau_{(x_0 \to D)}$ the time interval when the system **first** attains the specific number x_{i_1,j_1} of (i_1,j_1)-molecules, then what is the probability density function, $P_{(x_0,D)}(\tau)$, of this random variable $\tau_{(x_0 \to D)}$, and what is the k-th moment of $\tau_{(x_0 \to D)}$?

If we define the set of states *D* as absorbing states, which means that the system dynamics will stop as soon as the system reaches any member of the set *D*, then the system stopping time, starting from the state x_0, will be $\tau_{(x_0 \to D)}$. We will briefly introduce our methods and the results.

Final Performance Report

First, we introduce a new matrix S, which is obtained from the matrix Q by deleting all rows and columns corresponding to states in D. Secondly, we need to determine the number of states in D, denoted Dn, and introduce an $(N\text{-}Dn)$-dimensional row vector θ, which is 1 in the corresponding position of x_0 and zero otherwise. Based on the concept of phase type distribution introduced by Neuts (1982), we can derive the probability distribution of the eventual reaction time (ERT) by $P\left(\tau_{(x_0 \to D)} \leq t\right) = 1 - \theta e^{St} e$, $t \geq 0$, where e is a $(N\text{-}Dn)$-dimensional column vector with one in all elements.

Furthermore, by several mathematical arguments, we may now derive and summarize our main results as follows:

> The probability density function of the eventual reaction time (ERT) is

$$p_{(x_0,D)}(t) = -\theta e^{St} Se = \sum_{n=1}^{\infty} b_n^E(x_0) \frac{t^{n-1}}{(n-1)!}, \quad t \geq 0,$$

where e is a $(N\text{-}Dn)$-dimensional column vector with one in all elements, and $b_n^E(x_0)$ is given in terms of the propensity functions and state change vectors as follows:

$$b_1^E(x_0) = -a_0(x_0) + \sum_{k=1}^{M^E} a_k(x_0) = -a_0^D(x_0) \equiv -\sum_{k=1}^{M^D} a_k^D(x_0)$$

$$b_n^E(x_0) = -a_0(x_0) b_{n-1}^E(x_0) + \sum_{k=1}^{M^E} a_k(x_0) b_{n-1}^E(x_0 + v_k), \quad \text{for n} = 2, \cdots.$$

> The k-th moment of this eventual reaction time (ERT) is:

$$E[\tau_{(x_0 \to D)}]^k = (-1)^k k! \left(\theta S^{-k} e\right), \quad k = 1, 2, 3, \ldots.$$

> If we further ask "which reaction triggers this ERT event?" we have the following result. Let R_1 be the set of reactions such that after each reaction of R_1 a (i_1, j_1)-molecule is produced as a product. Therefore,

$$P\left(\text{reaction } R_j \text{ triggers this ERT event}\right) = \frac{c_j}{\sum_{k \in R^{(1)}} c_k}, \quad R_j \in R_1.$$

> If we denote by $P_{(x_0,D)}(t, j)$ the joint density function that the eventual reaction time is t and results as a consequence of reaction j, i.e.,

$$P\left(t \leq \tau_{(x_0 \to D)} \leq t + dt \text{ and reaction } R_j \text{ triggers this event}\right) = P_{(x_0,D)}(t, j) dt$$

then
$$P_{(x_0,D)}(t, j) = \left(\frac{c_j}{\sum_{k \in R^{(1)}} c_k}\right) \sum_{n=1}^{\infty} b_n^E(x_0) \frac{t^{n-1}}{(n-1)!} \quad \text{for } t \geq 0.$$

9

3.3.3. Case Study.

By using a simple reaction network, the isomerization reaction, The PI reached the following conclusion.

- Completed the numerical analysis based on both our novel analytical method and direct simulation analysis for the test reaction network.
- The results from numerical analysis and simulation analysis are almost exactly same, which supports the validity of our theoretical result.
- The time to calculate the solution of CME using our method is about 10 times faster than the time using the standard methods.
- Our method is expected to be superior, in terms of the running time, for more complex systems. This is currently being tested.

Final Performance Report

3.4 Matlab Code and the corresponding diagrams

A. The Matlab code for calculating the Network Chemical Master Equation (CME) by using the new proposed iterative method is as follows:

```
function PxT = GetCMEPxT()
%GETPXTS Summary of this function goes here
%   Detailed explanation goes here
C = .2;
X0 = (0:1:10);
X1 = (10:-1:0);
A1 = X0*C;
A2 = X1*C;
A0 = A1+A2;
N = 1000;
t = (0:.1:10);
curSta = 0; %Initial State PcurSta(x,t)
PxT = zeros(length(X0), length(t));
for tV = 1: length(t)
   for x = 1 : length(X0)
      %Get SumD
      SumD = GetDn(N, x, A1, A2, A0, t(tV));
      PxT(x,tV) = SumD(N, curSta+1); %Our Index start at 1 not 0 so we add 1
   end
end
plot(t, PxT);
xlabel('Time t');
ylabel(sprintf('P%d(x,t)', curSta));
title(sprintf('P%d(x,t)', curSta));
legend(sprintf('P%d(X0,t)',curSta), sprintf('P%d(X1,t)',curSta),
sprintf('P%d(X2,t)',curSta), sprintf('P%d(X3,t)',curSta), sprintf('P%d(X4,t)',curSta),
sprintf('P%d(X5,t)',curSta), sprintf('P%d(X6,t)',curSta), sprintf('P%d(X7,t)',curSta),
sprintf('P%d(X8,t)',curSta), sprintf('P%d(X9,t)',curSta), sprintf('P%d(X10,t)',curSta));

%filNam = sprintf('Result/PxT.xls');
%celNam = sprintf('PxT');
%xlswrite( filNam, PxT, celNam, 'A1')

function SumD = GetDn(N, y, A1, A2, A0, t)
MaxStates = length(A0);
D = zeros(N, MaxStates);
for n = 1: N
   for x = 1: MaxStates
      if n == 1
         if x == y
            D(n,x) = 1;
```
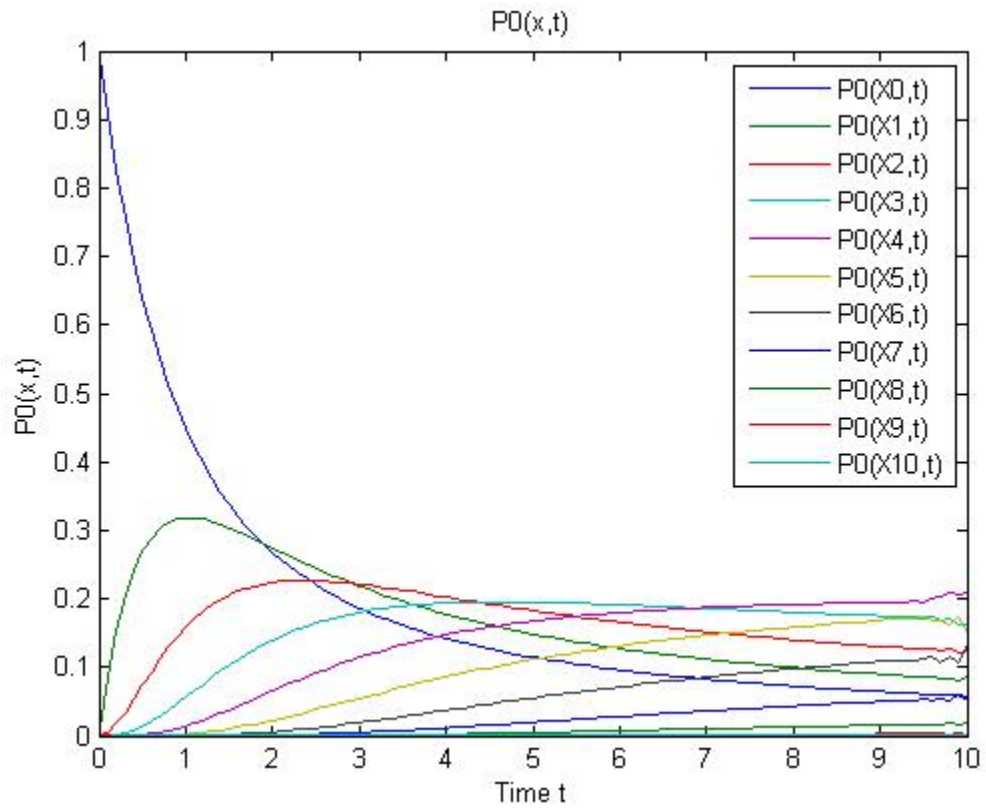11

```
        else
            D(n,x) = 0;
        end
      else
        D(n,x)= -A0(x)*D(n-1,x);
        if x > 1
            D(n,x) = D(n,x) + A1(x)*D(n-1, x-1);
        end
        if x < MaxStates
            D(n,x) = D(n,x) + A2(x)*D(n-1, x+1);
        end
        D(n,x) = D(n,x) * t / n;
      end
   end
end
SumD = zeros(N, MaxStates);
for n = 1: N
   for x = 1: MaxStates
     if n == 1
        SumD(n,x) = 0;
     else
        SumD(n,x) = SumD(n-1,x) + D(n,x);
     end
   end
end

for n = 1: N
   for x = 1: MaxStates
     if y == x
        SumD(n,x) = SumD(n,x) + 1;
     end
   end
end
```

Final Performance Report

The corresponding diagrams are:

PO(x,t)

Final Performance Report

B. The Matlab code for calculating the network Eventual Reaction Time (ERT) by using the new proposed iterative method is as follows:

```
function PxT = GetERTPxT()
%GETPXTS Summary of this function goes here
%   Detailed explanation goes here
C1 = .2;
C2 = 0.05;
X0 = (0:1:10);
X1 = (10:-1:0);
A1 = X0*C1;
A2 = X1*C2;
A0 = A1+A2;
N = 450;
t = (0:1:100);
curSta = 0; %Initial State PcurSta(x,t)
PxT = zeros(length(X0), length(t));
for tV = 1: length(t)
   for x = 1 : length(X0)
      %Get SumD
      SumD = GetCn(N, x, A1, A2, A0, t(tV));
      %PxT(x,tV) = SumD(N, curSta+1); %Our Index start at 1 not 0 so we add 1
   end
end
% plot(t, PxT);
plot(SumD);
% xlabel('Time t');
% ylabel(sprintf('P%d(x,t)', curSta));
% title(sprintf('P%d(x,t)', curSta));
% legend(sprintf('P%d(X0,t)',curSta), sprintf('P%d(X1,t)',curSta),
sprintf('P%d(X2,t)',curSta), sprintf('P%d(X3,t)',curSta), sprintf('P%d(X4,t)',curSta),
sprintf('P%d(X5,t)',curSta), sprintf('P%d(X6,t)',curSta), sprintf('P%d(X7,t)',curSta),
sprintf('P%d(X8,t)',curSta), sprintf('P%d(X9,t)',curSta), sprintf('P%d(X10,t)',curSta));

 %filNam = sprintf('Result/PxT.xls');
 %celNam = sprintf('PxT');
 %xlswrite( filNam, SumD, celNam, 'A1')

 function SumC = GetCn(N, y, A1, A2, A0, t)
MaxStates = length(A0);
C = zeros(N, MaxStates);
for n = 1: N
   for x = 1: MaxStates
      if n == 1
         if x == 9
            C(n,x) = t * (-A2(x));
```
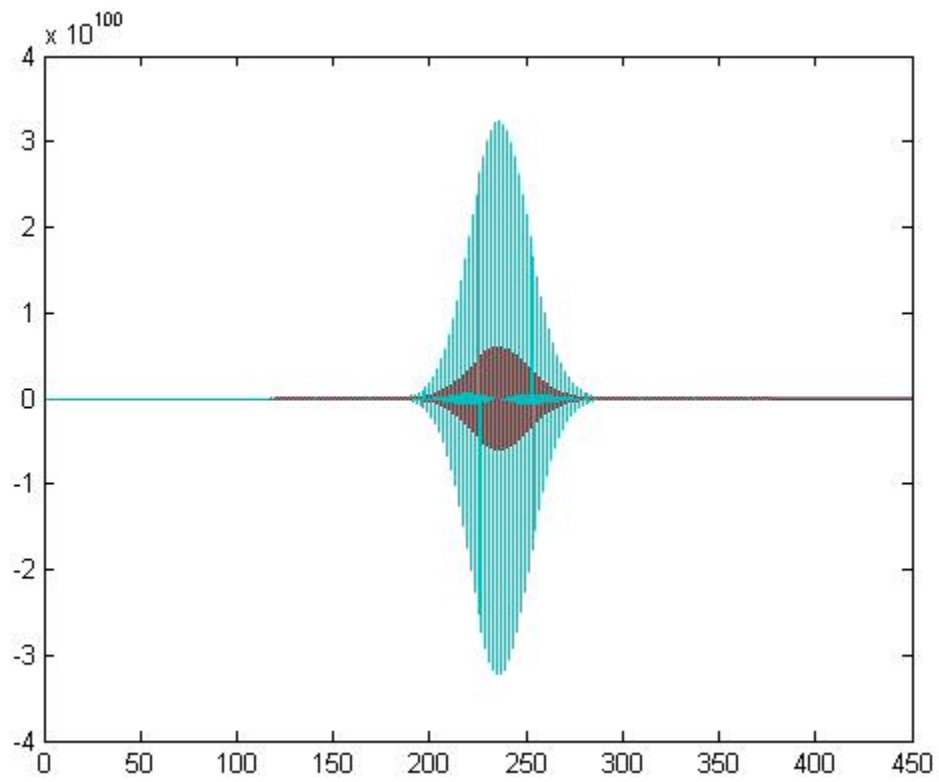
```
        else
            C(n,x) = 0;
        end
      else
        C(n,x)= -A0(x)*C(n-1,x);
        if x > 1
            C(n,x) = C(n,x) + A1(x)*C(n-1, x-1);
        end
        if x < MaxStates
            if x ~= 9
                C(n,x) = C(n,x) + A2(x)*C(n-1, x+1);
            end
        end
        C(n,x) = C(n,x) * t / n;
      end
   end
end
SumC = zeros(N, MaxStates);
for n = 1: N
   for x = 1: MaxStates
      if n == 1
        SumC(n,x) = 0;
      else
        SumC(n,x) = SumC(n-1,x) + C(n,x);
      end
   end
end
%
%for n = 1: N
%   for x = 1: MaxStates
%      if y == x
%         SumC(n,x) = SumC(n,x) + 1;
%      end
%   end
%end
```

Final Performance Report

The corresponding diagrams are:

4. **Conclusion**

The research of this project developed a new mathematical description of the biomolecular networks that control cellular behavior, formulate both novel analytic and computational tools to investigate the behavior of these networks and apply these approaches to evaluate the performance of engineered bio-constructs. The PI has investigated and developed a novel general stochastic biomolecular network description of cellular control systems from gene to output function. He provided a detailed framework to describe the components, reactions, and internal and external connections of these systems, built a probability model of the generalized stochastic biomolecular network, developed an analytic approach for the time-dependent solution and investigated computational methods to approximate these solutions. Specifically, the PI successfully applied the stochastic model and analytical methods to have answered specific questions about system behavior, which include define the "distance" between two nodes in terms of the time it takes for a variation introduced at the first node to propagate to the second node, investigated the hysteresis effect upon reversal of the nodes, and computed the distribution of time intervals from a given current state to a defined final state and identify the most likely reaction pathway that makes this happen. By successful dissemination of these technical advances, further technical progresses can be expected by collaborations and technical interactions with colleagues.

The project partially supported two Master students in the Department of Computer Science at Texas Southern University, one of the major Historically Black Colleges and Universities (HBCU) in the nation with high concentration of minority students from under-represented groups. The project directly impacted the capabilities, interests, and careers of the graduate students who participate in the project as well as any other students of interest.